

ノード間類似度を用いたオントロジーの 切り出し手法についての提案

板垣弦矢^{†1} 撫中達司^{†1} 齊藤志保^{‡2} 森郁海^{‡2} 折本拓真^{‡2} 伊藤岳広^{‡2}

概要: オントロジーはドメインにおける知識を明示的に表現するための仕様としてこれまで様々な分野で利用されてきた。中でも近年、行動認識の分野ではより高次の行動の抽出や行動の意図を理解するためにこれまで主流であったデータ駆動型に対して知識駆動型の行動認識が検討されている。一方で行動認識においてはユースケースによってプライバシーに関する情報を含む場合やネットワークに繋がりにくい環境での動作に対応する必要があるためローカルなエッジ端末上での動作を想定する必要がある。しかし、知識駆動型の既存研究の多くはオントロジーの利用方法や実装環境などの実用面での評価には至っておらず、例えば利用者が増加することを想定した場合オントロジーが拡大することになり、エッジ端末のようなリソースの限られた環境では実装できない可能性が考えられる。そこで本稿では拡大したオントロジーから語彙に関する予備知識なく包含的な切り出しを実現するノード間類似度を用いたオントロジーの切り出し手法について提案する。実験の結果、包含的な切り出しを実現することはできなかったが包含範囲付近まではクエリの意図を解釈できていたことから実現可能性の見込みを得た。

キーワード: オントロジー, RDF, 切り出し, 類似度, エッジ

1. はじめに

オントロジーはドメインにおける知識を明示的に表現可能にする仕様としてこれまで様々な分野で利用されてきた。中でも近年、行動認識の分野では従来のデータ駆動型から行動に関する知識を有したオントロジーを利用する知識駆動型が新たに提案されており、これにより高次の行動の抽出[1]や行動目的の理解 [2][3]といったことが検討されている。

一方、行動認識を用いるユースケースの中にはプライバシーに関する情報を含む場合やネットワークに繋がりにくい環境での動作を考慮する必要があるが、この様なユースケースに対してはエッジコンピューティングと呼ばれる処理手法により対応する方法がある。エッジコンピューティングはローカルなネットワークの出口付近にエッジ端末と呼ばれる処理能力を有した端末を設け、この端末上でプライバシーに関する処理や最低限の機能の提供などを可能としているが、エッジ端末はクラウドのような仮想環境と比べCPUやメモリ、ストレージなどのハードウェアスペックが必ずしも充分ではない。

前述した知識駆動型の研究の多くはオントロジーのスキームの定義や推論エンジンの提案は行っているが実用を想定したオントロジーの利用方法や実装環境については十分な評価を行っておらず、仮に多くの人々が当該オントロジーを利用する状況を想定した場合、多くの知識を記述する必要があり、オントロジーの拡大が想定される。しかしエッジ端末では通常複数のアプリケーションが動作していることから拡大したオントロジーの実装はリソースの圧迫に繋がる恐れがある。このことから拡大したオントロジーを

より実用的に利用するためには実装環境に合わせてオントロジーを対応させる事が必要であり例えばオントロジーの切り出しとその活用が重要である。

切り出しには RDF 形式のオントロジーを対象とした既存のキーワード検索手法が利用できるが、当該手法では必ずキーワードと同じ語彙がノードに含まれている必要があることからオントロジーに依存した語彙を予め知っておく必要があり、キーワードと同じ様な意味を持つ異なる語彙に広く対応することができない。

そこで本稿では、クエリとして記述される RDF トリプルとオントロジー内の RDF トリプルとの類似度を指標として切り出しを実行するノード間類似度を用いたオントロジーの切り出し手法について提案する。

2. 関連研究

オントロジーに多くの知識を記述することを前提とした提案の一つとして中村らの研究[1]があげられる。この研究ではライフログサービス向けに高次行動を抽出する基盤の構築を目的として低次行動や環境音、場所など様々な情報の組合せを高次行動として紐づけ記述するオントロジーを定義している。例えば、“映画を見ている”といった高次行動に対しては低次行動として“座っている”、“環境音として“音楽が流れている”、場所として“映画館にいる”という情報をそれぞれ紐付けることで表現している(図1)。しかし、“映画を見ている”という高次行動は例えば家にいながらも行えるため必ずしも場所を映画館と限定することはできず、また、低次行動についても寝ながら、立ちながら、歩きながらなど様々な状況が想定される。

^{†1} 東海大学 情報通信学研究科
Graduate School of Information and Telecommunication Engineering, Tokai University

^{‡2} 三菱電機株式会社 情報技術総合研究所
Mitsubishi Electric Corporation Information Technology R&D Center

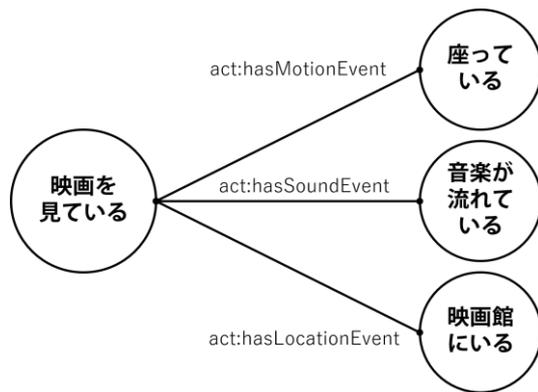


図 1 高次行動の抽出を目的としたオントロジーの記述例

このように高次行動を表現するための情報の組合せは無数に存在することから実用的なオントロジーを構築するためには情報の組合せを数多く記述する必要があり、これについては中村らも今後の課題としている。

一方で前述した中村らの研究を含め行動に関する知識を記述したオントロジーはパブリックに公開されていることが少なく、また網羅的に記述されているものは筆者の知る限りほとんどない。しかし、行動に属する語彙として解釈可能なドメインのオントロジーのひとつとしては国立研究開発法人農業・食品産業技術総合研究機構（農研機構）がパブリックに公開している農作業基本オントロジー（以下、AAO）[4]がある。AAOはAGROVOCと呼ばれる農林水産、食糧、環境、などの専門用語を32,000語以上網羅する世界標準のシソーラスデータや内閣府が農作物名称の標準化のために提案したガイドラインをもとに語彙の追加や階層構造の修正を施し構築されたオントロジーである[5]。目的と手段の2つの属性によって概念が大きく分類されており、それに対し場所や時期、対象作物といった様々な属性を紐付けることで具体的な情報を明確にしている。

拡大することが予想されるような前述のオントロジーやAAOは様々な要求に対応するため多くの知識を含んでいるが、個人が利用することを想定した場合、実際に必要となるのはその一部である。したがって多くの知識を含むオントロジーの中から特定の部分だけを抽出しエッジ端末向けのオントロジーとして新規にオントロジーを構築することができれば前述したユースケースにも対応できると考えられる。

オントロジーから特定の部分を抽出できる手法のひとつに伊藤の提案がある[6]。この提案では、ユーザ別に最適化されたWeb情報の提供を目的としており、Web情報に紐づく語彙をノードに設定した上でそれらをシナプスと呼ばれる重みを持ったノードで繋ぐオントロジーを構築している（図2）。シナプスの重みをWeb上でのユーザの挙動に合わせて更新することによりパーソナライズが行われ、ユーザがシステムを継続的に利用する中で重要と判断できる部分を半自動的に抽出すること可能としている。

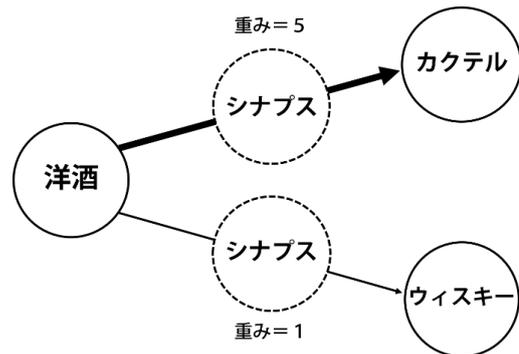


図 2 重みを用いたオントロジーのパーソナライズ

伊藤の提案を用いたオントロジーの切り出しについて考えると、既存オントロジーのRDFトリプルに対してシナプスを追加し、エッジ端末のリソースに合わせて重みを指標に数値の低い部分からノードを切り落としていく手法が考えられる。しかし、当該手法では重みの決定までに時間がかかることから一定期間、部分抽出を行うことができないという課題がある。

一方、オントロジーは一般的にRDF (Resource Description Framework) [7] と呼ばれるW3Cによって規格化された形式で記述されることが多く、RDFに対してはSPARQL[8]と呼ばれる専用のクエリ言語が標準化されている。このことから、SPARQLを用いることで重みのような蓄積された情報を必要としない部分抽出が可能であり、中でもSPARQLを用いたキーワード検索ではクエリ文法やドメインに関する知識が無くとも簡単にクエリを実行することが可能である。しかしSPARQLを用いたキーワード検索はメモリ効率や応答速度の改善が課題であったことからその解決策として浜松らのキーワード検索[9]が提案されている。この手法ではスキップ法を用いてキーワードを含むノードを抽出し、到達可能リストを用いてそれらのノード間距離を計算、最終的に任意のノード間距離を持つセットについてノードエッジ情報を補完し出力している（図3）。

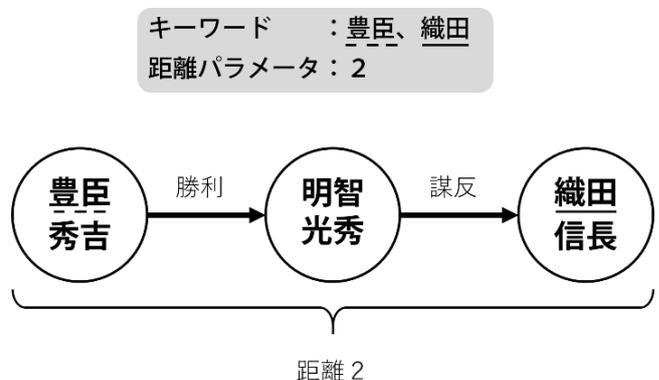


図 3 キーワード検索におけるオントロジーの抽出

浜松らの提案を用いたオントロジーの切り出しについて考えると、ノード間距離を指標としてエッジ端末のリソースに合わせて距離の近いものから RDF パスを複数選択し、それらを包含する上位ノードで結合する手法が考えられる。しかし、当該手法では必ずキーワードと同じ語彙がノードに含まれている必要があることからオントロジーに依存した語彙を予め知っておく必要があり、キーワードと同じ様な意味を持つ異なる語彙に広く対応することができない。

このことからオントロジーの切り出しにおいては語彙を予め知る必要のない語彙の概念的な解釈を用いた包含的な切り出しが課題となる。

以降、本稿の3章では課題解決のための提案手法について述べる。また、4章では実験結果について述べ、5章では結果の考察を述べる。

3. 提案手法

3.1 用語の定義

RDF は主語 (subject), 述語 (predicate), 目的語 (object) の3つの要素から構成され、これら要素をセット (s, p, o) にしたものを RDF トリプルと呼ぶ[10]。主語 s は URI 参照の集合 U , 空ノードの集合 B のいずれかの要素 $s \in U \cup B$ であり、述語 p は U の要素 $p \in U$, 目的語 o は U, B , リテラルの集合 L のいずれかの要素 $o \in U \cup B \cup L$ である。また、主語と目的語が一致する RDF トリプルを数珠繋ぎにした閉路を含まない $2d+1$ 個の要素列を $(r_0, p_1, r_1, p_2, r_2, p_3, \dots, r_{d-1}, p_d, r_d)$ と表現し RDF パスと呼ぶ。 r は主語と目的語が一致するリソースを指し d はパスの長さを指す。 RDF トリプルの有限集合 $G \in (U \cup B) \times U \times (U \cup B \cup L)$ を RDF グラフと呼ぶ。なお本稿におけるオントロジーは RDF グラフと同義である。

3.2 提案手法の概要

本稿では、クエリとして記述される RDF トリプル形式の基準トリプルとオントロジー内の RDF トリプルとの類似度を指標に切り出しを実行するノード間類似度を用いたオントロジーの切り出し手法について提案する (図 4)。本提案手法は RDF トリプルを形態素解析により単語の集合として扱い、既存の単語間類似度計算手法を用いた上でその結果を合算し RDF トリプル同士の類似度とする。これによりオントロジーに記述されている語彙を予め知る必要なく基準トリプルからオントロジー内の RDF トリプルを包含的に抽出することを可能としている。また抽出された RDF トリプルに対してルートノードまでのパスを補完することで結合を可能とし、包含的に抽出された RDF トリプルから新規にオントロジーを構築する。

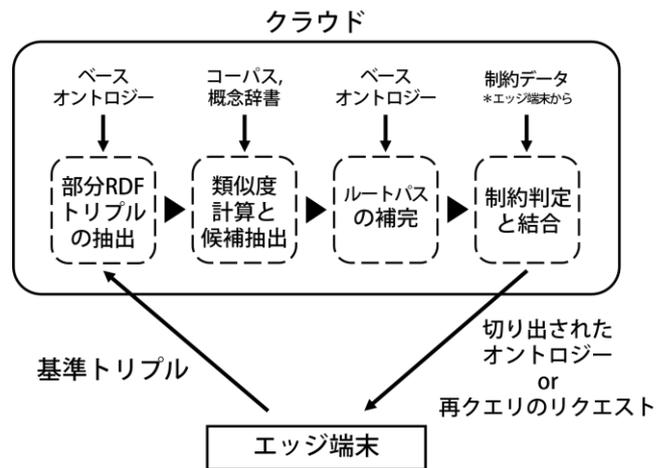


図 4 提案手法における処理の流れ

3.3 部分 RDF トリプルの抽出

類似度を算出するに当たり切り出し元となるオントロジー (以下、ベースオントロジー) 内の全 RDF トリプルに対して類似度計算を行うのは効率的ではない。そのためここでは類似度計算の前処理として基準トリプルのスキームを元に部分 RDF トリプルをベースオントロジーから抽出する。具体的に基準トリプルが

$s = \text{adl:作業}$
 $p = \text{property:*}$
 $o = \text{property:包丁}$

と発行された場合は、基準トリプルと同じスキームの関係を持つ部分 RDF トリプル $(\text{adl:}\sim, \text{property:}\sim, \text{property:}\sim)$ を図 5 のように抽出する。

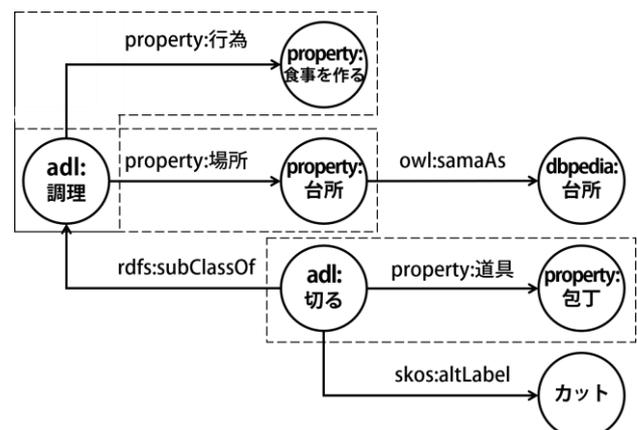


図 5 基準トリプルを元にした部分 RDF トリプルの抽出

このとき基準トリプルやベースオントロジーに記述されたスキーム以降の語彙部分については後述する類似度計算の対象であるため抽出処理には影響を及ぼさない。また、基準トリプルの中にリテラルを含む場合はその部分におけるスキームを限定しない。

3.4 類似度計算と候補抽出

基準トリプルと抽出した部分 RDF トリプルとの類似度を計算するにあたり、本手法では単語間類似度を扱う。そのため、双方のトリプルから類似度計算に必要な単語を抽出するためにスキーム以降の語彙部分もしくはリテラルを用いる。この時、基準トリプルについては単語が入力されることを想定する。しかし、ベースオントロジーから抽出した部分 RDF トリプルについては短文で記述されている場合が想定される。このことから本手法ではまず抽出した部分 RDF トリプルを 1 つ選択し、その語彙部分もしくはリテラルに対して形態素解析を行う。このとき主語、述語、目的語、別々に形態素の集合として取得する。次に得られた形態素の集合を主語は主語同士、述語は述語同士、目的語は目的語同士で基準トリプルとのすべての組み合わせに対して既存手法を用いた単語間類似度を計算する。そして最後に計算したすべての単語間類似度を合算することで基準トリプルと抽出した部分 RDF トリプルとの類似度とし、これら処理を抽出したすべての部分 RDF トリプルに対して行う。なお、単語間類似度の計算時、基準トリプルに特定の語彙を指定しない(ワイルドカードの*にあたる)ものがある場合は用いた単語間類似度の計算手法における最大の類似度を出力し、一方で語彙不足などにより類似度を計算できない場合は最低の類似度を出力する。また、得られた単語間類似度については単語の品詞によって異なる重みを設定することができ、これにより助詞や接続詞などの重要でない単語の影響を制御することができる。この重みはノードとエッジ、つまりは主語、目的語と述語で異なる値が設定できる。具体的な図及び数式を以下に示す(図 6)(式 1)。

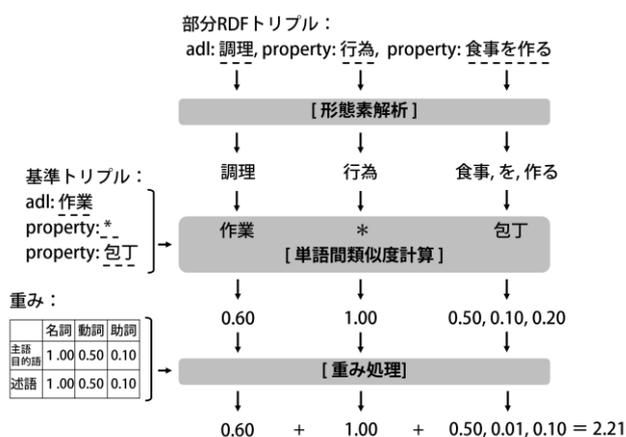


図 6 基準トリプルと部分 RDF トリプルとの類似度計算

$$\text{類似度} = \sum_{i=1}^I Wif(Ni) + \sum_{j=1}^J Wjf(Mj) \dots \text{(式 1)}$$

N は主語と目的語における形態素の和集合であり、 I はその要素数である。また、 Ni は和集合における特定の要素で

あり Wj はそれに対応する重みである。 M は述語における形態素の集合であり、 J はその要素数である。また、 Mj は述語集合における特定の要素であり Wj はそれに対応する重みである。 f は基準トリプルとの単語間類似度を算出する関数である。

類似度計算の完了後、最後に任意の閾値を満たす部分 RDF トリプルを切り出しの候補として抽出する。類似度の合算と候補抽出の具体例を 3.3 節の図 5 の内容をもとに表 1 に示す。候補抽出の閾値は仮に 2.40 とした。類似度を用いることでキーワード検索では対応できない“調理”や“切る”といった語彙を含むトリプルが切り出しの候補として抽出できていることが確認できる。

表 1 類似度の合算と候補抽出

部分 RDF トリプル	基準トリプル			合計類似度	閾値を満たす (閾値:2.40)
	行動	*	包丁		
1	0.60 (調理)	1.00 (行為)	0.50, 0.01, 0.10 (食事, を, 作る)	2.21	×
2	0.60 (調理)	1.00 (場所)	0.80 (台所)	2.40	○
3	0.60 (切る)	1.00 (道具)	1.00 (包丁)	2.60	○

3.5 ルートパスの補完

ここでは結合処理のための前処理としてすべての候補の RDF トリプルに対してベースオントロジーを用いて上位ノードを再帰的に代入しルートノードまでの RDF パスを補完する。具体的に以下の流れで補完を実行する(図 7)。

- ① 候補の RDF トリプルもしくは再帰処理の中で得られた RDF トリプルについて主語を目的語に配置したトリプルを発行しクエリを行う。
- ② ①で得られた RDF トリプルの中に述語として包含関係や上位関係を表すものがある場合それを①の RDF トリプルに対して代入する。
- ③ 包含関係や上位関係を表す述語が出現しなくなるまで ①, ②を繰り返す。

3.4 節の表 1 で候補として抽出された RDF トリプルについては図 7 のように補完される

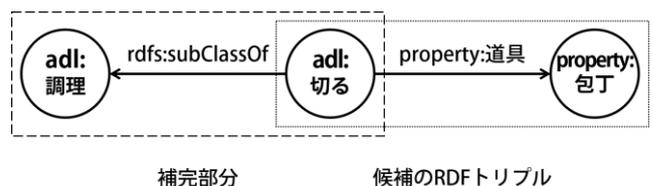


図 7 RDF トリプルのルートパス補完

なお、本手法におけるルートパス補完では閉路を含まない木構造のベースオントロジーを想定している。

3.6 制約判定と結合

エッジ端末の余剰リソースなどを元に設定される制約データを用いてこれまでの処理で得られた RDF パスについてその最大距離やノード数などから制約判定をする。制約条件を満たさなかった場合はエッジ端末に対して再クエリのリクエストを行い、満たした場合はすべての RDF パスをルートノードで結合しエッジ端末へ出力する (図 8)。

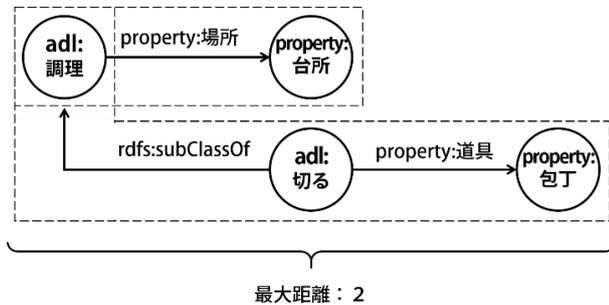


図 8 RDF パスの結合

4. 実験

4.1 実装

本稿では行動認識分野におけるオントロジーの拡大を切り出しの大きなモチベーションとしていることから本来は行動認識に関するオントロジーを用いた実験を行うべきである。しかし、前述の通りパブリックに公開されている行動に関するオントロジーの中には切り出しの効果を十分に検証できるだけの知識を持ったオントロジーを見つけることはできなかった。そこで今回は行動に属する語彙として解釈が可能なドメインのオントロジーであれば本来の行動に関するオントロジーと同等の結果が得られると考え、農作業に関する幅広い知識を持つ AAO を用いる。

単語間類似度の計算手法についてはこれまで様々な提案がされているが、本稿における課題が語彙の概念的な解釈を用いることから単語の意味的解釈が可能な計算手法として分類されるコーパスベースと知識ベースの[11]のものを対象として検討した。それぞれ代表的なものとしては word2vec[12]や WordNet[13]を用いる LIN[14]などが挙げられる。WordNet を用いる手法については語彙の上位下位関係などを厳密に解釈できるが例えば、喜びと怒りといった全く反対のイメージを持つ言葉を感情という上位概念で繋げていることから一般的な感覚とは異なる類似度を算出する可能性がある。そのため今回は日本語 Wikipedia の全文データから作られたコーパス[15]を元に word2vec を用いて類似度を算出する。

なお、単語を抽出するための形態素解析ツールには Janome を使用しており、品詞の重みはノードとエッジどちらも名詞が 1, その他の全て品詞は 0.2 に設定した。

4.2 実験方針とクエリ (キーワード) の決定

提案手法評価のためキーワード検索との比較検証を行う。それにあたり出力される RDF パスにおいて類似度や単語間距離の良いものから上位 10 個を比較対象とし、類似度計算の効果을明らかにするためクエリ (キーワード) に設定する語彙についてはオントロジー内に含む場合と含まない場合の 2 パターンで検証する。

今回の実験では農家がナスの育て方に関する部分を切り出すことを想定し、表 2 のようにクエリ (キーワード) を設定した。

表 2 実験のためのクエリ (キーワード)

手法	オントロジーに語彙を含む	オントロジーに語彙を含まない
キーワード検索	繁殖, ナス	栽培, 茄子
提案手法	繁殖, 作物, ナス	栽培, 作物, 茄子

4.3 結果

4.2 節の表 2 で設定したクエリ (キーワード) の実験結果を以下に示す (図 9)。なお、詳細内容については付録 1 ~ 付録 3 に記載する。

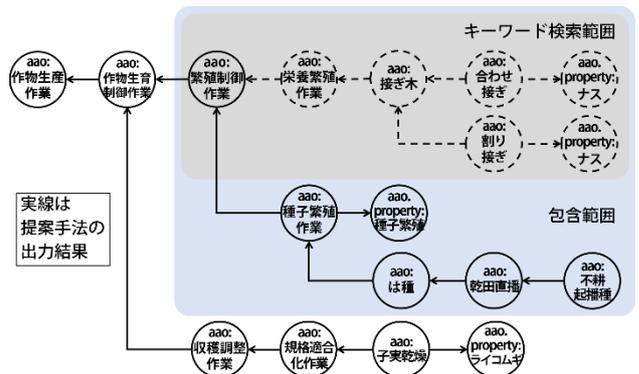


図 9 実験結果の概略図

キーワード検索の抽出した RDF パス集合をキーワード検索範囲とし、提案手法における包含的な切り出しの範囲 (包含範囲) をキーワード検索範囲に含まれるノードとそれを頂点とする配下の RDF パスすべてとして比較する。

オントロジー内に語彙を含まない場合のキーワード検索に関しては手法上 RDF パスを出力することができない。一方、オントロジー内に語彙を含んだ場合のキーワード検索に関しては“繁殖制御作業”から“ナス”までの RDF パスを出力することができた。これに対しオントロジー内に語彙を含んだ場合の提案手法では 10 個の出力結果のうち 2 つは“繁殖制御作業”配下のノードを含んだ RDF パスであったことから包含範囲内であったがそれ以外はすべて包含範囲外であった。また、オントロジー内に語彙を踏まない場合の提案手法ではすべての RDF パスが包含範囲外であ

り、本稿の課題であった語彙を知る必要のない包含的なオントロジーの切り出しを実現できていない結果となった。

5. 考察

提案手法が課題を解決できなかった原因を考察し改善方法について述べる。まず、提案手法の出力に影響する要素として処理順にクエリ、オントロジー、形態素解析、単語間類似度計算、単語間類似度の合算の5つが挙げられる。中でも形態素解析以降の要素については提案手法の核であり、一方クエリやオントロジーはあくまでも入力であることから今回は形態素解析以降の要素について重点的に考察を行う。

まず、本提案における形態素解析の役割は単語間類似度計算のためにオントロジーから入力される語彙部分を単語分割することである。この単語分割には通常、辞書が用いられている。しかしこの辞書は用いる形態素解析ツールによって内容が異なることから、特に複合名詞に対しては単語分割の程度がツールによって変わる。この時、単語間類似度計算に使われているコーパスが形態素解析の出力する単語を語彙として保有しているかが大きな問題となる。この問題に対してはコーパス作成時に用いる形態素解析ツールと単語間類似度計算前に用いる形態素解析ツールを統一することで単語分割の違いをなくすることができる。本実験において用いた形態素解析ツールは Janome であるが Janome は辞書として MeCab のデフォルトシステム辞書を利用している。一方コーパスに使用した日本語 Wikipedia の全文データは同じく MeCab で作成されていることから本実験において当該問題は発生していないと考える。

次に単語間類似度計算について考察する。前述の通り単語間類似度計算において重要な問題は利用しているコーパスが入力された単語を語彙として保有していないことである。本実験におけるコーパスの語彙不足について調べたところ 295 単語が語彙不足であったために類似度計算されていなかった。これについては AAO に専門用語が多く含まれていることが原因であると考えられ、この語彙不足により本来、類似度の高い RDF トリプルが候補として抽出されなかった可能性がある。実際、キーワード検索の出力した RDF パスに含まれる“接ぎ木”、“挿し木”、“割り”、“接ぎ”、“栄養”などの単語は語彙不足のため類似度計算されていない。一方、提案手法ではクエリでナス（茄子）と入力していたにもかかわらずライコムギなどの穀物類の語彙を含む RDF パスが複数出力されていた。これについてもコーパス依存の問題があると考え、出力される単語間類似度を調査するため、実際に AAO に含まれる農作物とナスとの単語間類似度を計算した。結果、ライコムギやイネの単語間類似度が分類としてより近いと考えられるトマトやピーマンなどの野菜より高いことがわかった（表 3）。

表 3 ナスと他農作物との類似度

農作物	類似度
ナス	1
ライコムギ	0.886369
キュウリ	0.74236
イネ	0.740081
トマト	0.675696
サツマイモ	0.662637
ピーマン	0.616249

これらコーパス依存の問題に対してはドメイン特有の単語や語彙の繋がりをコーパスに学習させることで改善できる可能性があり、また、AAO の様に上位下位関係を厳密に分類しているオントロジーに対しては同様に厳密な概念分類を行っている WordNet を用いた単語間類似度計算がより相性が良い可能性がある。

最後に単語間類似度の合算について考察する。単語間類似度の合算では品詞の重みを調整することでノイズの軽減を図ることができるが名詞については単語間類似度の合算において重要な品詞であることから通常は重みを高く設定することが考えられる。しかし、コーパスが複合名詞として認識できない場合、単語分割により名詞の形態素が多くなってしまいうため語彙部分が長いものと短いものでは類似度に大きく差が出てしまうことが想定される。このような場合、単語間類似度を合算するのではなく最大値や平均値を取ることで正規化ができる。そこで実際に栽培、作物、茄子のクエリに対して最大値で類似度計算を行い RDF トリプルの候補を抽出、RDF パスまでの補完を行った。出力結果の詳細については付録 4 に記載するが、“植物ホルモン剤による受粉”のような語彙部分が極端に長いものについては除外することができた。しかし、新たに包含範囲内の RDF パスが追加されることはなく、前述した穀物類の語彙を含む RDF パスについては除外できていない。このことから本実験においてはコーパスに依存した語彙不足の影響が大きいと考えられる。

一方で、提案手法の出力した全ての RDF パスにおいてルートノードから作物生産作業までは同じパスをたどっており、キーワード検索の最上位ノードであった制御繁殖作業についても同様に作物生産作業を通っていることを考慮すると包含範囲付近まではクエリの意図を解釈することはできていることからコーパスの改善による包含的なオントロジーの切り出しの可能性が見込まれる。しかしコーパスの改善は人手で実施され、多くの費用や時間がかかることが想定されることからキーワード検索と提案手法の双方の特徴を組合せた解決についても合わせて検討が必要である。

6. まとめ

予め語彙を知る必要のない包含的なオントロジーの切り出しを課題としてノード間類似度を用いたオントロジーの切り出し手法を提案した。実験の結果、課題の解決には至らなかったが包含範囲付近まではクエリの意図を解釈することができていたことから実現可能性の見込みを得た。今後の課題は単語間類似度計算におけるコーパスの改善やキーワード検索と提案手法の特徴を組合せた解決の検討である。

謝辞 今回の実験にて使用した農作業基本オントロジーの内容につき、大変丁寧かつ詳しい説明をいただいた農研機構、竹崎氏に感謝する。

参考文献

- [1] 中村亮, 石川佳治, 杉浦健人, 脇田佑希子. “ライフログサービスのためのオントロジーに基づく行動イベント処理”. DEIM Forum, 15-1, 2017.
- [2] Chen Liming, Nugent Chris D, Wang Hui. “A knowledge-driven approach to activity recognition in smart homes”. IEEE TKDE, Vol 24, Issue 6, pages 961-974, 2012.
- [3] Rafferty Joseph, Nugent Chris D, Liu Jun, Chen Liming. “From Activity Recognition to Intention Recognition for Assisted Living Within Smart Homes”. IEEE Transactions on Human-Machine Systems, Vol 47, Issue 3, pages 368-379, 2017.
- [4] 共通農作業語彙 (CAVOC), CAVOC (オンライン), 入手先 <<http://www.cavoc.org/aa0.php>> (参照 2019-07-23).
- [5] 朱成敏, 武田英明, 法隆大輔, 竹崎あかね, 吉田智一. “農業 IT システム間データ連携のための農作業基本オントロジーの構築”. 信学技報, vol.115, no. 290, 149-154, 2015.
- [6] 伊藤浩. “オントロジーを利用した情報パーソナライズ, カスタマイズ方法の提案”. SIG-SWO-A404, 03, 2005.
- [7] RDF 1.1 Concepts and Abstract Syntax, W3C Recommendation (オンライン), 入手先 <<https://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/>> (参照 2019-07-23).
- [8] SPARQL 1.1 Overview, W3C Recommendation (オンライン), 入手先 <<https://www.w3.org/TR/2013/REC-sparql11-overview-20130321/>> (参照 2019-07-23).
- [9] 浜松良樹, 兼岩憲. “RDF グラフに対するキーワード検索の高速化と省メモリ化”. SIG-SWO-042, 05, 2017.
- [10] 兼岩憲. “RDF と RDF スキーマの推論”. 人工知能学会論文誌, Vol. 26, No. 5, pp. 473-481, 2011.
- [11] Wael H. Gomaa, Aly A. Fahmy. “A Survey of Text Similarity Approaches”. International Journal of Computer Applications. Volume 68, No.13, April 2013.
- [12] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, Jeffrey Dean. “Distributed Representations of Words and Phrases and their Compositionality”. NIPS. 2013.
- [13] George A. Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross, Katherine Miller. “Introduction to WordNet: An On-line Lexical Database”. International Journal of Lexicography. Volume 3, Issue 4, Pages 235-244, 1990.
- [14] Lin D. “An information-theoretic definition of similarity”. Proceedings of the 15th International Conference on Machine Learning. pp.296-304, 1998.
- [15] GitHub - Kyubyong/wordvectors: Pre-trained word vectors of 30+ languages, GitHub (オンライン), 入手先 <<https://github.com/Kyubyong/wordvectors>> (参照 2019-07-23).

付録

付録 1 ~ 付録 4 内の “>” は `rdfs:subClassOf` の関係を指しており, 各行末端にあるカンマつなぎの 3 つ組は RDF トリプルを表している. なお, 各語彙に付随するスキームについては省略している.

付録 1 キーワード検索×語彙を含む (繁殖, ナス)

No.	出力結果 (RDF パス)
1	栄養繁殖作業>接ぎ木> 合わせ接ぎ, 対象作物, ナス
2	栄養繁殖作業>接ぎ木> 割り接ぎ, 対象作物, ナス
3	栄養繁殖作業>接ぎ木> 挿し接ぎ, 対象作物, ナス
4	栄養繁殖作業>接ぎ木>合わせ接ぎ> 断根合わせ接ぎ, 対象作物, ナス
5	栄養繁殖作業>接ぎ木>挿し接ぎ> 断根挿し接ぎ, 対象作物, ナス
6	繁殖制御作業>栄養繁殖作業>接ぎ木> 合わせ接ぎ, 対象作物, ナス
7	繁殖制御作業>栄養繁殖作業>接ぎ木> 割り接ぎ, 対象作物, ナス
8	繁殖制御作業>栄養繁殖作業>接ぎ木> 挿し接ぎ, 対象作物, ナス
9	繁殖制御作業>栄養繁殖作業>接ぎ木>合わせ接ぎ> 断根合わせ接ぎ, 対象作物, ナス
10	繁殖制御作業>栄養繁殖作業>接ぎ木>挿し接ぎ> 断根挿し接ぎ, 対象作物, ナス

付録 2 提案手法×語彙を含む (繁殖, 作物, ナス)

No.	出力結果 (RDF パス)
1	農作業>基本農作業>作物生産作業>収穫調製作業 >収穫作業>収穫物集約作業> 飼料作物梱包, 対象作物, 飼料作物
2	農作業>基本農作業>作物生産作業>作物生育制御 作業>繁殖制御作業> 種子繁殖作業, 目的, 種子繁殖
3	農作業>基本農作業>作物生産作業>作物生育制御 作業>生殖成長制御作業>結実制御作業>受粉> 化学的手段による受粉> 植物ホルモン剤による受粉, 機資材, 植物ホルモ ン剤
4	農作業>基本農作業>作物生産作業>収穫調製作業 >規格適合作業> 子実乾燥, 対象作物, ライコムギ

5	農作業>基本農作業>作物生産作業>作物生産環境制御作業>土壌制御作業>土壌成分制御作業>施肥>追肥> 実肥散布, 対象作物, ライコムギ
6	農作業>基本農作業>作物生産作業>作物生育制御作業>繁殖制御作業>種子繁殖作業>は種>乾田直播> 不耕起播種, 対象作物, イネ
7	農作業>基本農作業>作物生産作業>作物生産支援作業>有用生物飼育作業> アイガモ飼育, 対象作物, イネ
8	農作業>基本農作業>作物生産作業>作物生産支援作業> 有用生物飼育作業, 目的, 有用生物飼育
9	農作業>基本農作業>作物生産作業>作物生産環境制御作業>土壌制御作業>土壌成分制御作業>施肥>追肥> 穂肥散布, 対象作物, ライコムギ
10	農作業>基本農作業>作物生産作業>作物生育制御作業>栄養成長制御作業>順化作業>根回し> 幼植物ずらし, 対象作物, 野菜

付録3 提案手法×語彙を含まない(栽培, 作物, 茄子)

No.	出力結果 (RDF パス)
1	農作業>基本農作業>作物生産作業>収穫調製作業>収穫作業>収穫物集約作業> 飼料作物梱包, 対象作物, 飼料作物
2	農作業>基本農作業>作物生産作業>作物生産環境制御作業>土壌制御作業>土壌成分制御作業>施肥>追肥> 実肥散布, 対象作物, ライコムギ
3	農作業>基本農作業>作物生産作業>作物生産環境制御作業>土壌制御作業>土壌成分制御作業>施肥>追肥> 穂肥散布, 対象作物, ライコムギ
4	農作業>基本農作業>作物生産作業>収穫調製作業>規格適合化作業> 子実乾燥, 対象作物, ライコムギ
5	農作業>基本農作業>作物生産作業>作物生産支援作業>機資材管理作業>機資材設置作業>誘引資材設置> 果樹棚設置, 対象作物, 果樹
6	農作業>基本農作業>作物生産作業>作物生産支援作業>機資材管理作業>機資材調整作業> 果樹棚補修, 対象作物, 果樹
7	農作業>基本農作業>作物生産作業>作物生産支援作業>運搬作業>商品運搬> 玄米運搬, 対象作物, イネ

8	農作業>基本農作業>作物生産作業>作物生育制御作業>生殖成長制御作業>結実制御作業>受粉>化学的手段による受粉> 植物ホルモン剤による受粉, 機資材, 植物ホルモン剤
9	農作業>基本農作業>作物生産作業>作物生産環境制御作業, 目的, 作物生産環境制御
10	農作業>基本農作業>作物生産作業>作物生育制御作業>生殖成長制御作業>結実制御作業>無核化>化学的手段による無核化> 植物ホルモン剤による無核化, 対象作物, 果樹

付録4 栽培, 作物, 茄子(クエリ)の最大値候補抽出

No.	出力結果 (RDF パス)
1	農作業>基本農作業>作物生産作業>収穫調製作業>収穫作業>収穫物集約作業> 飼料作物梱包, 対象作物, 飼料作物
2	農作業>基本農作業>作物生産作業>作物生産支援作業>有用生物飼育作業> アイガモ飼育, 対象作物, イネ
3	農作業>基本農作業>作物生産作業>作物生育制御作業>栄養成長制御作業>順化作業>根回し> 幼植物ずらし, 対象作物, 野菜
4	農作業>基本農作業>作物生産作業>収穫調製作業>規格適合化作業> 子実乾燥, 対象作物, オオムギ
5	農作業>基本農作業>作物生産作業>収穫調製作業>規格適合化作業> 子実乾燥, 対象作物, コムギ
6	農作業>基本農作業>作物生産作業>収穫調製作業>規格適合化作業> 子実乾燥, 対象作物, ライコムギ
7	農作業>基本農作業>作物生産作業>収穫調製作業>規格適合化作業> 子実乾燥, 対象作物, エンバク
8	農作業>基本農作業>作物生産作業>収穫調製作業>規格適合化作業> 子実乾燥, 対象作物, ライムギ
9	農作業>基本農作業>作物生産作業>収穫調製作業>規格適合化作業> 子実乾燥, 対象作物, イネ
10	農作業>基本農作業>作物生産作業>作物生産支援作業>機資材管理作業>機資材調整作業> 果樹棚補修, 対象作物, 果樹